

**PATENT APPLICATION**

**HIGH SPEED DATA ACCESS MEMORY ARRAYS**

**Inventor:** Luigi Di Gregorio, a citizen of Belgium, residing at  
321 Coty Way  
San Jose, CA 95136

**Assignee:** Telairity Semiconductor, Inc.  
3375 Scott Boulevard, Suite 300,  
Santa Clara, CA, 95054

**Entity:** Small

TOWNSEND and TOWNSEND and CREW LLP  
Two Embarcadero Center, 8<sup>th</sup> Floor  
San Francisco, California 94111-3834  
Tel: 650-326-2400

## HIGH SPEED DATA ACCESS MEMORY ARRAYS

### BACKGROUND OF THE INVENTION

[0001] The present invention relates to memory arrays, and more particularly to high speed techniques for reading data from memory arrays.

[0002] Prior art memory arrays includes rows and columns of static random access memory cells (SRAM). Each SRAM cell includes two cross coupled inverter circuits that store one bit of data. In a multi-port SRAM, each row of memory cells is coupled to a read word line and a write word line, and each column of memory cells is coupled to a read bit line and a write bit line.

[0003] In a single port SRAM, the read word line and the write word line for each row of memory cells are the same. Also, the read bit line and write bit line for each column of memory cells are the same in a single port SRAM.

[0004] By activating the write word line for a selected row, a data bit can be written into a selected memory cell in the selected row. The data bit is transmitted to the selected memory cell along the write bit line that controls the selected memory cell.

[0005] By activating the read word line for a selected row, a data bit can be read from a selected memory cell in the selected row. The data bit is transmitted from the selected memory cell outside the memory array along the read bit line that controls the selected memory cell.

[0006] Once a data bit has been read from a memory cell, the data bit can be buffered and amplified using a differential sense amplifier. Differential sense amplifiers are used to reduce the parasitic effects of noise on output signals from a memory array. Noise can cause a bit to be read incorrectly. For example, noise can cause a logic high to be interpreted as a logic low. Differential amplifiers cancel out the effects of noise on output signals from a memory array.

[0007] Sometimes the bit line capacitance is very high. When the bit line capacitance is high, the differential voltage is slow to develop across the differential bit lines. A differential amplifier amplifies the differential bit line signal and results in a faster access time for the memory.

[0008] Some prior art memory arrays use inverter circuits to sense the bit line voltage. Because voltages on the bit lines change slowly (due to high capacitances), an inverter is slow to sense the bit line voltage.

5 [0009] The output signals of the inverters are multiplexed by a multiplexer. The multiplexer selects a signal from one of the inverter circuits.

[0010] Typically, a single multiplexer in a memory array multiplexes signal from numerous read bit lines. For example, a single multiplexer may multiplex signals from 8 bit lines. Prior art multiplexers are slow to read bits from a memory array, because they have to multiplex signals from many read bit lines. Also, long wires may extend from the outputs of the

10 inverters to the inputs of the multiplexer, further slowing down read time.

[0011] Therefore, it would be desirable to provide faster techniques for reading data bits from memory cells in a memory array.

#### BRIEF SUMMARY OF THE INVENTION

15 [0012] The present invention provides high speed techniques for reading data from memory arrays. Memory arrays comprise rows and columns of individual memory cells. Bits are accessed from the memory cells along local read bit lines.

20 [0013] The local read bit lines are coupled to logic gates. The input terminals of each logic gate are coupled to receive signals from two of the local read bit lines. Bits on the local bit lines are passed to the logic gates. In one embodiment of the present invention, the logic gates are NAND gates.

[0014] The logic gates are coupled to a global bit line. Pull down transistors cause the voltage on the global bit line to change state when the output signal of one of the logic gates changes state.

25 [0015] Resistive and capacitive (RC) delays on the global bit line become a problem when the length of the wire becomes too long. Memory arrays of the present invention can have multiple global bit lines that are isolated from each other to reduce RC delays. Repeater circuits are coupled in between two global bit lines so that a signal on one global bit line can propagate to another global bit line.

30 [0016] The techniques of the present invention apply to single ended and differential read bit lines. For differential bit lines, the techniques of the present invention can be applied on each one of the differential bit lines.

[0017] Other objects, features, and advantages of the present invention will become apparent upon consideration of the following detailed description and the accompanying drawings, in which like reference designations represent like features throughout the figures.

5

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0018] FIG. 1 illustrates a multi-port SRAM cell;

[0019] FIG. 2 illustrates two local read bit lines that are coupled to a NAND gate according to an embodiment of the present invention;

[0020] FIG. 3 illustrates two NAND gates and pull down transistors that couple four local

10 read bit lines to a global bit line according to an embodiment of the present invention;

[0021] FIG. 4 illustrates a NAND gate that is driven by two global bit lines according to an embodiment of the present invention;

[0022] FIG. 5 illustrates a transistor level schematic of circuitry that couples two local bit lines to a global bit line according to an embodiment of the present invention; and

15 [0023] FIG. 6 illustrates a first global bit line that drives a second global bit line through a repeater circuit according to an embodiment of the present invention.

#### DETAILED DESCRIPTION OF THE INVENTION

[0024] The present invention provides high speed techniques for reading data bits from memory arrays. Memory arrays comprise numerous memory cells that are usually arranged in rows and columns. Memory arrays of the present invention can comprise SRAM cells, dynamic random access memory (DRAM) cells, read only memory (ROM) cells, electrically erasable programmable read only memory (EEPROM) cells, EPROM cells, programmable logic arrays (PLAs), FLASH memory, and many other types of memory cells.

[0025] An example of a multi-port SRAM cell that can be used in a memory array of the present invention is illustrated in FIG. 1. The SRAM cell 110 of FIG. 1 includes two cross coupled inverters coupled between supply voltage  $V_{DD}$  and ground. The first inverter includes p-channel transistor 104 and n-channel transistor 105. The second inverter includes p-channel transistor 106 and n-channel transistor 107.

[0026] SRAM cell 110 is controlled by the write word line WWL0. When the signal on write word line WWL0 goes HIGH, n-channel transistor 111 is turned ON, coupling SRAM cell 110 to write bit line WBL0. A data bit can be written into SRAM cell 110 from write bit line WBL0 through transistor 111.

[0027] Write bit line WBL0 can write a HIGH or LOW bit into SRAM cell 110.

Transistors 108 and 109 couple a second input of SRAM cell 110 to ground when the signals on write lines WWL0 and WBL0 are HIGH.

[0028] When the signal on read word line RWL0 goes HIGH, transistor 101 turns ON.

5 When transistor 101 is ON, read bit line RBL0 is coupled to SRAM cell 110 through the inverter that includes transistors 102 and 103. When the signal on RWL0 goes high, a data bit is read out of SRAM cell 110 along read bit line RBL0. Inverter 102/103 buffers and inverts output signals from SRAM cell 110. RBL0 is always precharged before a READ occurs.

10 [0029] A first embodiment of the present invention is illustrated in FIG. 2. Sixteen n-channel MOS field-effect transistors 202-209 and 221-228 are shown in FIG. 2. Transistors 202-209 are coupled to read bit line 230, and transistors 221-228 are coupled to read bit line 231.

15 [0030] Each one of transistors 202-209 couples read bit line 230 to a corresponding memory cell. Also, each one of transistors 221-228 couples read bit line 231 to a corresponding memory cell. For example, transistor 101 in FIG. 1 is an example of transistors 202-209 and 221-228. The corresponding memory cell may be an SRAM cell as shown, for example, in FIG. 1. The sources of transistors 202-209 and 221-228 are shown as being coupled to ground merely to simplify the diagram.

20 [0031] Transistors 202-209 are turned ON and OFF by read word lines rwl0a-rwl7a, respectively. Transistors 221-228 are turned ON and OFF by read word lines rwl0b-rwl7b, respectively. Only one read word line (RWL) can be turned on at a time.

25 [0032] Read bit line 230 is pre-charged to supply voltage  $V_{DD}$  through p-channel transistor 201. When  $V_{PRE}$  goes to GND (i.e., ground), transistor 201 turns ON, and bit line 230 is pre-charged to near  $V_{DD}$ . Read line 231 is pre-charged to  $V_{DD}$  through p-channel transistor 229. When  $V_{PRE}$  goes to GND, transistors 229 turns ON and, and bit line 231 is pre-charged to near  $V_{DD}$ .

30 [0033] Read bit lines 230 and 231 are both coupled to inputs of NAND gate 210 as shown in FIG. 2. When a signal on one of read word lines rwl0a-rwl7a or rwl0b-rwl7b, goes HIGH, a corresponding one of transistors 202-209 or 221-228 turns ON. A bit stored in a memory cell is passed to read bit line 230 or 231 through the transistor 202-209 or 221-228 that is ON.

[0034] If a LOW bit is transferred to a read bit line 230 or 231, the read bit line is pulled LOW through one of transistors 202-209 or 221-228.  $V_{PRE}$  is HIGH so that transistors 201 and 229 are OFF. When one of the input signals at NAND gate 210 goes LOW, the output

signal  $V_{RO}$  of NAND gate 210 goes HIGH. Only one of read bit lines 230-231 goes LOW at a time. When a HIGH bit is transferred to read bit line 230 or 231, the output signal  $V_{RO}$  of NAND gate 210 remains LOW.

[0035] NAND gate 210 replaces the multiplexer of the prior art. NAND gate 210 passes a signal from one of two read bit lines to its output. Signals pass through NAND gate 210 faster than a traditional multiplexer. Prior art multiplexers are typically implemented as an inverter (on a sense bit line) followed by a transmission gate and another inverter (i.e., a multiplexer). The present invention multiplexes the read bit lines with only a one gate delay. Also, in the present invention, NAND gate 210 passes a signal from a read bit line to its output without having to wait for a select signal.

[0036] FIG. 3 illustrates a further embodiment of the present invention. Signals on read bit lines are transmitted to NAND gate 210 as discussed above. The output of NAND gate 210 is coupled to n-channel transistor 311. Transistor 311 is coupled between global bit line 330 and ground.

[0037] Memory cells are also coupled to other local read bit lines in the memory array. For example, memory cells are coupled to local read bit lines 321 and 322 shown in FIG. 3. Read bit lines 321 and 322 are also pre-charged to near the supply voltage. Read bit lines 321-322 are coupled to NAND gate 320.

[0038] When a signal on one of read bit lines 230-231 goes LOW, the output of NAND gate 210 goes LOW, and transistor 311 turns ON. Global bit line 330 is initially pre-charged to near the supply voltage  $V_{DD}$  through p-channel transistor 310. When transistor 311 turns ON and transistor 310 is OFF, the voltage on global bit line 330 is pulled LOW.

[0039] When a signal on one of read bit lines 321-322 goes LOW, the output of NAND gate 320 goes LOW, and transistor 312 turns ON. When transistor 312 turns ON and transistor 310 is OFF, the voltage on global bit line 330 is pulled LOW.

[0040] Bits are transmitted from memory cells in the memory array to global bit line 330 without the use of a traditional multiplexer. Global bit line 330 is isolated from the local read bit lines 230, 231, 321, and 322. Therefore, the small capacitance and resistance on the local read bit lines does not slow down signals transmitted on global bit line 330.

[0041] Global bit line 330 is longer than the local bit lines. Therefore, global bit line 330 has a much larger capacitance and resistance than the local bit lines. For example, global bit line 330 can be 1 millimeter long. Global bit line 330 can be as long as the wire RC delays permit timing requirements to be met.

[0042] Pull down transistors 311 and 312 have much larger channel widths than transistors 202-209 and 221-228. Because global bit line 330 is a long wire (i.e., with a high capacitance), transistors 311-312 need to draw more current to pull down the voltage on global bit line 330 and to provide a fast discharge. The pull down transistors provide a

5 significant increase in speed over prior art techniques that had a long first level bit line and no second level bit line.

[0043] Each of the local bit lines can be coupled to any suitable number of memory cells depending on speed requirements. For example, local bit line 230 can connect to 8, 16, or 32 memory cells. Global bit line 330 can be coupled to any suitable number of pull down

10 transistors. For example, global bit line 330 can be coupled to 2-16 pull down transistors. Only one bit from a memory cell at a time can be transmitted along a global bit line.

[0044] According to the present invention, bits from the memory cells are read using one local bit line for each column of memory cells. Prior art techniques used redundant, complimentary read bits lines for each column to drive differential amplifiers. The memory arrays of the present invention use less circuitry and are faster than prior art techniques. The techniques of the present invention can be used with single ended and differential read bit lines. For differential bit lines, the techniques of the present invention can be applied on each one of the differential bit lines.

[0045] In a further embodiment of the present invention, two global bit lines can be coupled to inputs of a second level NAND gate. This NAND gate performs the same function as NAND gate 210. Signals pass from the global bit lines through the second level NAND gate much faster than through prior art multiplexers that have many inputs (typically many more than 2).

[0046] An example of this embodiment is illustrated in FIG. 4. Global bit line 330 is coupled to a first input of NAND gate 450. Global bit line 430 is coupled to a second input of NAND gate 450. NAND gate 450 can drive signals from global bit lines 330 and 430 to a third level read bit line to provide further RC isolation. RC isolation speeds up signals passing on the global and third level bit lines by reducing the resistance and capacitance (RC) on these wires.

[0047] NAND gate 450 can also drive signals from the global bit lines to an output pin. Because NAND gate 450 is small device (e.g., four transistors) with only two inputs, it can quickly drive signals from global bit lines 330 and 430 to its output.

[0048] Global bit line 430 is pre-charged HIGH by pull up transistor 425 and pulled down LOW by transistor 421. Transistor 421 has a large channel width so that it can provide a fast discharge of line 430.

[0049] Bits are transmitted from the memory cells along local read bit lines 410 and 411 to 5 first level NAND gate 420, as discussed above with respect to NAND gate 210. NAND gate 420 drives pull down transistor 421.

[0050] FIG. 5 illustrates a transistor level embodiment of pre-charge circuitry and a NAND gate of the present invention. The pre-charge circuitry of FIG. 5 pre-charges the local and global bit lines to a HIGH voltage during a pre-charge cycle. A pre-charge cycle is initiated 10 before a bit is read from a memory cell. A signal indicative of a stored bit can be passed from a memory cell by pulling one of the pre-charged local bits lines and the pre-charged global bit line LOW.

[0051] A first pre-charge voltage  $V_1$  is applied to an input of the pre-charge circuitry as 15 shown in FIG. 5. Voltage  $V_1$  is increased to a HIGH voltage to initiate a pre-charge cycle. P-channel transistor 501 and n-channel transistor 502 form a CMOS inverter. The output of inverter 501/502 is coupled to the gates of p-channel transistors 201 and 229. When  $V_1$  goes HIGH, the output voltage  $V_2$  of inverter 501/502 goes LOW, and transistors 201 and 229 turn ON.

[0052] Transistors 511 and 201 are coupled in parallel between high supply voltage  $V_{DD}$  20 and local bit line 230. When transistor 201 is ON, the voltage on local bit line 230 is pre-charged to a HIGH voltage. Transistors 513 and 229 are coupled in parallel between high supply voltage  $V_{DD}$  and local bit line 231. When transistor 229 is ON, the voltage on local bit line 231 is pre-charged to a HIGH voltage.

[0053] The output of inverter 501-502 is also coupled to the input of buffer 515. Buffer 25 515 includes two CMOS inverters. Buffer 515 delays and buffers the output signal of inverter 501/502. The output signal  $V_{PRE}$  of buffer 515 is applied to the gate of p-channel transistor 310. The delay of buffer 515 is preferably set so that transistors 310 and 311 are not ON at the same time.

[0054] After voltage  $V_1$  goes HIGH,  $V_{PRE}$  transitions LOW. Transistor 310 turns ON to 30 pre-charge global bit line 330. The output signal  $V_{PRE}$  of buffer 515 is also applied to the gates of other pre-charge transistors in the memory array.

[0055] Thus, when  $V_1$  goes HIGH to initiate a pre-charge cycle, the local and the global bit lines are pre-charged HIGH. When the pre-charge cycle ends,  $V_1$  goes LOW. A bit from a

memory cell can now be transmitted along the local and global bit lines as an output signal of the memory array.

[0056] Two p-channel transistors 521-522 and two n-channel transistors 523-524 form NAND gate 210 as shown in FIG. 5. Local bit line 230 is coupled to the gates of transistors 522 and 523, and local bit line 231 is coupled to the gates of transistors 521 and 524.

[0057] The output of NAND gate 210 is coupled to the gates of p-channel transistors 511 and 513. When the output of NAND gate 210 is LOW, transistor 511 keeps local bit line 230 at a HIGH voltage, and transistor 513 keeps local bit line 231 at a HIGH voltage. Transistors 511 and 513 are called “keepers,” because they keep the bit lines HIGH if they are not discharged during a read cycle. The keepers are very small p-channel transistors.

[0058] As long as bit lines 230-231 are both HIGH, the output of NAND gate 210 remains LOW, because transistors 523-524 are ON. When the voltage on local bit line 230 or local bit line 231 is pulled LOW, one of transistors 521-522 turns ON, pulling the gate of transistor 311 HIGH. Transistor 311 turns ON, pulling the voltage on global bit line 330 LOW. The LOW signal on the local bit line is transmitted to global bit line 330 in this manner.

[0059] FIG. 6 illustrates an embodiment of repeater circuits that drive signals from one global bit line to another global bit line. Referring to FIG. 6, read bit lines 230 and 231 are coupled to NAND gate 210. Read bit lines 321 and 322 are coupled to NAND gate 320. Read bit lines 631 and 632 are coupled to NAND gate 621. Read bit lines 633 and 634 are coupled to NAND gate 623.

[0060] N-channel transistors 311 and 312 pull down pre-charged global bit line 330. N-channel transistors 622 and 624 pull down a second global bit line 650. Global bit line 650 is pre-charged HIGH by p-channel transistor 641.

[0061] When transistor 311 or 312 turns ON, the voltage on global bit line 330 is pulled LOW. The LOW signal on bit line 330 can propagate to a second global bit line through repeater circuits 611 and 612. After the voltage on global bit line 330 goes LOW, the output of inverter 611 goes HIGH. Transistor 612 turns ON, pulling the pre-charged voltage on global bit line 650 LOW. Inverters 611 and 613 can be standard CMOS inverters.

[0062] The LOW signal on global bit line 650 can propagate to a third bit line through repeater circuits 613 and 614 in the same manner. When the voltage on line 650 goes LOW, the output of inverter 613 goes HIGH. When the output of inverter 613 goes HIGH, transistor 614 turns ON, pulling the pre-charged voltage on a third global bit line LOW.

[0063] Signals can propagate from one global bit line to another using the repeater circuitry shown in FIG. 6. The circuitry shown in FIG. 6 is merely one embodiment of the present

invention. Other types of repeater circuits can also be used to propagate signals from one bit line to another.

**[0064]** In addition, local read bit lines 631-634 can transmit bits from memory cells to global bit line 650 through circuits 621-624. Signals on segmented global bit line 650 that

5 originate from local bits lines 631-634 can propagate to other segmented global bits lines through repeater circuits 613 and 614.

**[0065]** Rather than providing one global bit line with a large capacitance and resistance, multiple segmented global bit lines are provided in the embodiment of FIG. 6. Capacitance and resistance on a wire tends to slow down signals propagating through the wire.

10 **[0066]** Segmented global bit lines 330 and 650 are shorter and therefore have less capacitance and resistance than one long global bit line. Repeater circuits 611 and 612 isolate the segmented global bit lines. Signals propagate faster through shorter global bit lines that are isolated by repeater circuits.

**[0067]** Signals originating from a memory cell can propagate to as many segmented global 15 bit lines as needed. FIG. 6 shows four local bits lines for each segmented global bit line. This is merely an example that is not intended to limit the scope of the present invention. Any suitable number of local bits can drive one global bit line.

**[0068]** While the present invention has been described herein with reference to particular embodiments thereof, a latitude of modification, various changes, and substitutions are

20 intended in the present invention. In some instances, features of the invention can be employed without a corresponding use of other features, without departing from the scope of the invention as set forth. Therefore, many modifications may be made to adapt a particular configuration or method disclosed, without departing from the essential scope and spirit of the present invention. It is intended that the invention not be limited to the particular 25 embodiment disclosed, but that the invention will include all embodiments and equivalents falling within the scope of the claims.